DOCUMENT RESUME

ED 241 550 TM 840 030

AUTHOR Boldt, Robert F.
TITLE Review for Perceived Bias on ASVAB Forms 11, 12, and 13.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RM-83-4
PUB DATE Dec 83
NOTE 18p.
AVAILABLE FROM Educational Testing Service, Research Publications, R-116, Princeton, NJ 08541.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Aptitude Tests; Attitudes; *Military Personnel; Opinions; *Test Bias; Testing Problems; Test Items; *Vocational Aptitude
IDENTIFIERS *Armed Services Vocational Aptitude Battery

ABSTRACT

The project reported here consisted of a sensitivity review of the items of Forms 11, 12, and 13 of the Armed Services Vocational Aptitude Battery (ASVAB). Because administration of this battery is a required step in the accession process, it should be free from perceived bias or offensiveness that could detract from the measurement process. In the percent study, two types of sensitivity reviews were conducted: one by Educational Testing Service (ETS) sensitivity reviewers and one by inductees during their first few days of Basic Training. This report summarizes the procedures and general results. In order to avoid compromising test security, specific item content is not reported here. Instead, examples and analogies are used to convey the kinds of responses encountered. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

RM-83-4

ED241550

RESEARCH MEMORANDUM

REVIEW FOR PERCEIVED BIAS ON ASVAB

FORMS 11, 12, AND 13

Robert F. Boldt

December 1983

TM 840 030

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
H. Weidenmiller

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Educational Testing Service
Princeton, New Jersey

Copyright © 1983. Educational Testing Service. All rights reserved.

REVIEW FOR PERCEIVED BIAS IN ASVAB

FORMS 11, 12, AND 13

INTRODUCTION

The project reported here consisted of a sensitivity review of the items of Forms 11, 12, and 13 of the Armed Services Vocational Aptitude Battery (ASVAB). Because administration of this battery is a required step in the accession process, it provides one of the first military experiences of enlistees. As such it should be free from perceived bias or offensiveness that could detract from the measurement process and set the enlistee off on the wrong foot. The knowledge gained in this project will help insure that the ASVAB will indeed be as free as possible from perceived bias or offensiveness.

Test taking is approached by the examinees in part on the basis of their backgrounds. Many of those taking the ASVAB are members of groups that are protected under the Civil Rights Act of 1964, and as such are experienced in trying to perform successfully under conditions that are particularly adapted to the White male group that has traditionally predominated in many vocational settings. They may become distracted by subtle aspects of the test situation, such as inadvertant implicit messages in the test items. For example, the consistent occurrance of males or Whites in dominant roles may convey to some female or minority examinees that they are not expected to do well on the test, nor perhaps in the service itself. The item writers do not intend to convey this message--the problem is in the inference rather than the implication.

Like the Department of Defense, Educational Testing Service(ETS) is concerned with the sensitivities of examinees. Because of this concern ETS has developed review processes that usefully applied to the ASVAB in this project. Our examinations are very much in the public view, and we have no wish to convey negative messages to potential examinees or to anyone else. Rather, our products must be free of irrelevant distractions so that the examinees are able to concentrate on demonstrating their potential when taking ETS tests. To this end we have for several years conducted sensitivity reviews of tests in our major programs. The review process consists of an examination of the items by ETS professionals, including members of protected groups, who are experienced at reviewing items. Where possible sensitivity issues arise, the items are discussed and modified until the offending aspects are eliminated without modifying the intellectual problem posed to the examinee. In the example above, a simple remedy is to put females and minority group members in some dominant positions. If successful modification cannot be accomplished, the item is scrapped, but seldom is an item lost in this way.

Note that the emphasis here is on perceived bias, as opposed to statistical bias. The term "perceived" is used to emphasize the fact that the study focuses on subjective judgements of the possible effects of items on special groups of people, and does not deal with item statistics. ETS also conducts other statistical studies of group differences in item performance in connection with its own test programs. Indeed, research on group differences is and has been one

of the research priorities of the organization. But the conduct of these studies is separate from the sensitivity, or perceived bias, review of the type with which this project is concerned.

In the present study, two types of sensitivity reviews were conducted: one by ETS sensitivity reviewers and one by inductees during their the first few days of Basic Training. The trainees are used because of our concern about the reactions of those who take the ASVAB under operational conditions; by testing trainees we get as near as is feasible to questioning the actual ASVAB testing population under operational conditions. Indeed, given that we will obtain the reactions of trainees to the tests, one might wonder why the review by ETS staff is necessary. There are two reasons for this. One is that the ETS staff will screen the material, eliminating those subtests for which trainee review seems unnecessary. In this way we are able to get the most useful information from trainees with the most efficient use of their training time. The second reason for an ETS staff review is that there are others, such as lawyers and staff, who might be inclined and in a position to raise criticism on the grounds of perceived bias, and who are better represented by ETS staff than by trainees. The review by ETS staff provides material in response to such criticism because it constitutes a rigorous review to detect subtle bias as well as to detect offensiveness.

The specific and detailed results of this project have been submitted in the form of a Special Appendix to appropriate client personnel. This report summarizes the procedures and general results. In order to avoid compromising test security, specific item content is not reported here. Instead, examples and analogies are used to convey the kinds of responses encountered.

ETS STAFF REVIEW FOR APPROPRIATE CONTENT AND LANGUAGE

For each form of the ASVAB there are two subforms of three of the ten subtests, which are combined into six batteries, any one of which might be administered to a particular examinee. All six of these were provided to each staff reviewer. The reviewers were also provided with a copy of the test evaluation criteria (Hunter & Slaughter,1980), which provide for examination of "individual test items and test(s) as a whole to...(detect) language, symbols, or content that are generally considered potentially offensive, inappropriate for major subgroups of the test-taking population, or serving to perpetuate any negative attitude that may be conveyed to these subgroups. No item in any test should include words, phrases, or descriptions that are generally regarded as biased, sexist, or racist(e.g. demeaning modifiers or stereotypes)." After reviewing all of the forms each reviewer provided written notes to the Principal Investigator. The six reviewers included three females, two male Blacks, and one male Hispanic.

Before proceding with the detailed comments, which focus on the problems noted, it should be mentioned that neither the forms examined in the present study nor the preceding three forms, which were also reviewed by ETS, contained problems that implied blatant disregard for groups of examinees. However, one of the reviewers of Forms 8, 9, and 10 noted many instances where women were presented less favorably than men. That same reviewer, when reporting on Forms 11,12, and 13 noted that the presentation of women was much improved. Reviewers oral comments on the new forms to the Principal Investigator were essentially favorable.

In the normal course of events at ETS the test development specialist and the reviewer would attempt to modify the test item so that its essential measurement characteristics could be retained without including the possibly sensitive aspects of the item. At ETS the philosophy of review is that if ANY reviewer sees a possible problem, it is treated as being there. This philosophy can be adopted because almost all items can be successfully revised, thus avoiding the loss of otherwise acceptable and hard-to-come-by test items. In contrast to the ETS procedure, the present project, which concerns non-ETS tests, terminates with a written report of the sensitivity analyses, which are given below.

General Science(GS)

Three of the General Science items posed sensitivity problems. Two of these used "man" to denote "humanity." One used the phrase "senses of man," which could be changed to "human senses," and the other used "man," which could be replace by "human beings." Modifying these items so that females are not excluded would have no other effect on the content of the items. The other item dealt with a biological condition related to skin color, and would not be saved in the course of events at ETS. As a matter of practice, items that refer to skin color, however innocuously, are not used at ETS simply because they are not essential for most purposes. In this case, it is difficult to see how familiarity with albinism is an essential portion of the science knowledge needed for performance in the service. Some other item could probably be substituted.

Arithmetic Reasoning(AR)

One problem occurring in the Arithmetic Reasoning items was an imbalance of male and female actors. In all cases the balance could be restored by replacing male actors with female actors, or by making the items gender neutral. There was one item where a jeweler was unnecessarily assumed to be male. Although this did not produce any strong reactions, it should be pointed out that tacit assumptions on the part of item writers can lead to imbalances of sex roles that leave women in subservient positions. However, the imbalances noted in the subtests examined here were not pejorative.

One item dealt with carpeting and padding, concepts that are not as familiar to those recently from Puerto Rico and other tropical areas.

A mixture item concerning meats used pork, which is forbidden to orthodox Jews and Muslims. Any possible problem could be avoided by using lamb as an ingredient of the mixture.

Word Knowledge(WK)

This subtest also contained an imbalance of sex references. Forms 11a, 12b and 13b were the most uneven in this regard. The problem could be remedied by changing the reference in some of the items or by making the items gender neutral.

Form 13b had a special problem in that the women were "luminous," "spotlessly clean," and "pinching affectionately." One reviewer commented that it would be better to give them a serious occupation.

Several of the items contained words that might be loaded for Blacks, although in the format used in Word Knowledge items the impact of the words might be quite attenuated. This format presents the word with no contextual material,

thus minimizing the provocative quality of the item display. However, there was an additional item, one to which Jewish examinees might be sensitive, that all the reviewers agreed is not an essential item in one's military vocabulary and to which many Jews have distinctly unpleasant associations. Because we cannot compromise the test item we will not give the test word here, but it would be similar to using the word "pogrom" with Jewish examinees. One of the alternatives in the items added to the problem, as would "desirable esthetic quality" as an alternative because to suggest that the combination of "pogrom" with "desirable esthetic quality" is conceivably an effective distractor is to demonstrate insensitivity to the possible problems of Jewish examinees. This and the General Science item on the biological skin condition are the only two items that would not be saved at ETS.

A different kind of possible problem arises because the examination may be taken by those for whom English is a second language. The difficulty occurs when the stem word is homophonically or homographically similar in the examinee's first and second languages, but the range of alternate meanings in the two langauges does not coincide, and the alternatives include one that is close to one of the meanings in the first language. Again we are unable to give the actual item, but the following example shows the principle. The word "actual" means, in Spanish, "presently." That is, for the examinee with a Spanish language background, "actual" is much more closely tied to the present time than it is to any notion of reality or truth. Thus an item with "actually" as the stem and "presently" and "really" as alternatives would give those most oriented to English an advantage because they would never think of confusing "actually" with "presently." Yet the error made by the Hispanic is not an error of misunderstanding or even of ignorance, and is fairly easily corrected.

Paragraph Comprehension(PC)

The example in the directions posed a possible problem for the reviewers. The stem of the example refers to the duties of the lighthouse keeper and their purpose in one sentence, and in the following sentence uses the masculine pronoun "he," allowing the examinees to make the incorrect inference that lighthouse keepers must be male, or supporting an inclination to believe that. More generally the use of the masculine pronoun in this context supports the common and unnecessary association of particular sexes with particular jobs or job roles and, unfortunately, does it in the context of this early military experience--the taking of the test. The problem is easily avoided by changing the phrase "even though he would like to aid in the rescue of its crew and passengers," to "though wishing to aid in the rescue of its crew and passengers."

The example poses a second problem in that there are two believable alternatives. The examinee has to choose whether, in the event of a nearby shipwreck, the lighthouse keeper should stay by the light to sound the siren. It might be possible to sound the siren without putting the light out or leaving the lighthouse, and if so, sounding the siren might be a good thing to do. But the alternative identified as correct is to stay at the light. The reviewers suggested that one of the alternatives should be modified. When modifications are made, the correct alternative, which is "stay at HIS light," should be changed to "stay at THE light"[emphasis is mine], for the same reasons as those given in the paragraph above.

The example in the directions contains unnecessary gender references that are easily corrected. The Paragraph Comprehension items also contained such references and used "man" for "humanity" as well, which is also easily corrected.

The reader is no doubt quite aware that certain groups have preferences about the ways in which they are identified. One of these is native Americans, many of whom resent the term "Indian." That term appeared in one of the test items, and one of the reviewers suggested that by using a specific identifier such as "Pueblo," which is not sensitive, the difficulty could be avoided without changing the intellectual problem one wishes the examinee to confront.

Another difficulty arises from the use of the words "America" and "American." There are north, south, and central Americans who are not citizens of the U.S.A. Also, there are citizens of the U.S.A. who live in the country but do not live on the North American continent (excluding Alaska). Some Hispanics dislike the use of "American" when what is meant is "citizen of the United States" because it seems to imply that the rest of the continent as well as South and Central America is of no consequence. The problem is easly corrected by saying, "In the United States six out of ten people..." rather than saying "Six Americans out of ten..." For passages that refer to the continental United States, one can simply write "continental United States," thus avoiding the acceptance of a tacit convention of excluding all other areas from the country unless they are explicitly included. Alaska poses something of a problem here, but we are not aware of any sensitivity problem concerning Alaska. The main problem is probably with Hispanics, for whom the methods suggested above should work.

Several of the reviewers identified the stem and alternatives of one of the items as potentially offensive to Hispanics. It dealt with the rise of cities in Latin America. It emphasized poor economic conditions and expressed a pessimistic attitude toward the urban developments. The tone of the passage was critical. The telling criticism of the stem and alternatives is that essentially the same material could be written without mentioning Latin America--the setting could be another place and time. Hence the identification of Latin American as the place with which the item was concerned could be regarded by sensitive Hispanics as a gratuitous slur.

Finally, one stem dealt with cross-country skiing, which was referred to as "inexpensive." The reviewers noted that this reflected middle-class standards that may not be shared by the many of the trainees.

Numerical Operations(NO)

Numerical Operations, which is a series of simple arithmetic computations, offers no problems. There were no comments.

Coding Speed(CS)

Like Numerical Operations, Coding Speed is also essentially neutral, but since a little verbal material is used there is the possibility for some sensitivity. In this subtest there is a key that associates words with numbers--for each word there is a number. There is a different key on each page. The item stems consist of the words and the alternatives consist of five of the numbers, one of which is associated in the key with the word in the stem. The examinee has to concentrate, going as swiftly as possible, on finding the number that goes with the word in the stem. Because that concentration is necessary,

"buzz words" should probably not appear in the keys and stems. "Jig," for example, is a word with several meanings, one of which is highly pejorative to Blacks. "Jig" probably should not be used as a stem word in the Coding Speed subtest even though it has several quite innocuous meanings. It did not actually appear in the ASVAB, but a similar word did. Words with homosexual coloring should probably be avoided as well. Omitting such words puts no hardship on the test developer because there are many short and clearly innocuous words that could be used and because there is no necessary coverage of content.

Auto and Shop Information(ASI)

The Auto and Shop items deal with material that is probably emphasized more for men than for women in our society, but within the subtest there is very little variation in the sensitivity of the material. Women may feel at a disadvantage when confronting this type of material, but the reviewers felt that the material had face validity for some jobs.

Mathematics Knowledge(MK)

The Mathematics Knowledge subtest differs from Arithmetic Reasoning in that the items of the former deal with highly abstract material. There are no human actors in the stems and alternatives. No sensitivity-related comments were made by the reviewers.

Mechanical Comprehension(MC)

The Mechanical Knowledge items are also almost entirely abstract and without human actors. There were, however, a few minor gender problems. The actors in these items were almost all identifiable as male, and unnecessarily so. In its publicity the Service associates women with formerly male jobs. Perhaps that could be carried out here as well.

Electronics(EI)

As with Mathematics Knowledge and Auto and Shop Information items the Electronics items are abstract and without human actors. Also as with these subjects, electronics is emphasized more for males than for females in our society, but the items appear to have face validity for some jobs. No sensitivity-related comments were made by the reviewers.

TRAINEE REVIEW FOR UNFAIRNESS OR OFFENSIVENESS

At the beginning of this project it was intended to choose a few subtests, all forms of which would be examined by the trainees. This course of action seemed feasible since examination of certain subtests of the ASVAB, such as Numerical Operations or Mathematics Knowledge, reveals essentially no sensitive material, as was noted by the ETS staff sensitivity reviewers. Feasibility of limiting the number of subtests examined also obtained because, when material was identified as sensitive on several forms of a subtest, the type of sensitivity, such as unnecessary gender specification, was repeated. Hence we could note the effects of unnecessary gender specification without requiring examination of all forms by the trainees. Furthermore, limiting the subtests presented to those for which potential sensitivity problems have been noted makes a more efficient use of the trainees' time. It was decided, therefore, to limit the subtests to be presented the using comments of the sensitivity reviewers. The resulting

allocation of subtests to trainee groups was as follows:

Group Number	Subtests and Form Examined
I	GS(11), AR(11a), PC(11a)
II	AR(11b), PC(11b), MC(11)
III	AR(12a), WK(11b), PC(12a)
IV	WK(12b), PC(12b), CS
V	GS(13), AR(13a), PC(13a)
VI	AR(13b), WK(13b), PC13b)

Note that not all of the subtest designations above are the same. There are six forms of AR, WK, and PC that are used, but not in all possible combinations. Use of six of the possible combinations leads to the battery designations of 11a, 11b, 12a, 12b, 13a, and 13b. There are only three forms of GS and MC. CS has no form designation because the CS material examined by the trainees consisted of pages taken from several of the forms, including those with stimuli that the sensitivity reviewers regarded as sensitive. This was done because it was not felt that examination of CS material by more than one group was necessary. The group number simply refers to the order in which the groups of trainees arrived, and there is no special significance to that order.

The trainees who participated in this study were in the first few days of induction processing at Fort Dix, New Jersey. Six trainee groups of approximately thirty each were requested with the groups to be mixed with respect to race and sex. No special attempt to obtain random samples was made because the study is much too small to support generalization on a statistical basis. The main thrust of the request to the installation was to provide groups that were mixed. Of the 215 trainees provided, 171 were male, 43 were females, and one failed to provide sex identification. The Blacks numbered 49, Hispanics numbered 16, one participant failed to provide race identification. The remaining 149 classified themselves as "other." Base personnel were careful to ensure that the groups were mixed by race and sex as requested by ETS. To have more Hispanic participants would have been preferable if they had been available.

Instructions to the trainees were to indicate whether the items were fair and spoke of people respectfully. The words "unfair" and "inoffensive" were used several times in the instructions. The trainees were to circle the numbers of the items that had sensitivity problems, indicate on the booklet the group to which each item might be unfair, and mark the part of the item that caused the trouble. Appendix A contains the directions that were in the trainees' booklets as well as an approximate script of the session administrators' remarks. One session had a Black administrator with a White female proctor. In the other session the administrator was Hispanic with a Black proctor, both male. The results of the trainees' sessions are described below.

In this discussion there appear at the end of the paragraphs evaluative comments that often use the editorial "we." These comments are based on comments made to the Principal Investigator by ETS sensitivity staff and trainees.

General Comments

By far the most salient result was that the trainees had very little to say about the subtests. It is possible that they were overwhelmed by their new life situation and felt uncomfortable about being asked to criticize the tests.

But they expressed interest in the fact that we were asking their opinions about the sensitive aspects of the tests, and in the fact that the military establishment had a concern about such matters. The tone of the sessions was positive, and we noticed no strong overt expression of dissatisfaction with the tests.

Because gender imbalance and the unnecessary use of gender were frequently noted by the ETS reviewers, the general trainee reaction to these problems was of special interest. However, we found no evidence that gender references are upsetting to examinees even though we attempted to maintain an extremely permissive atmosphere. Only one comment was made by one person on these matters. Gender imbalance might not have gotten a completely fair test in this trial since we did not stress sensitivity problems posed by combinations of items. However, the purpose of the sessions was not to raise consciousness; if the identification of gender imbalance as a sensitive issue only arises by focussing on it, then it is clearly not blatently offensive.

Though the trainees' comments were scattered, there were some consistencies that might be noted. Many comments were concerned with identifying groups that might be expected to perform relatively less well on some of the items becuase they may have had less exposure them. Items that dealt with automobiles were thought to be unfair to people who hadn't had cars, and problems about brokers were thought to be unfair to people who did not deal with the stock market. One trainee commented that the schools in one midwestern city taught very elementary science courses, and that students from minority groups did not take them in any case. But clearly some items must be written in a context, and any context can concievably put some group at a disadvantage. The test constructor must simply provide a reasonably wide coverage of contexts so that the advantages and disadvantages can be fairly spread. We are not critical of the tests on this ground.

Group I

This group consisted of 33 trainees, 27 of whom had no comment on the booklet. None of the six who made comments on the items made them on the same item. That is, no item received more than one comment. One of the people, a male, made the only comment about an unnecessary gender reference. The item noted concerned a woman. Since none of the eight females in the group reacted to this item, the reference does not seem to engender sensitivity. The group contained 25 males and eight females, seven Blacks and two Hispanics. The only Jewish trainee in the study was in this group.

Group II

This group consisted of 37 trainees, eighteen of whom had no comment.

The Arithmetic Reasoning items in this book drew only two comments from more than one trainee, both of which were commented on by two persons. One of these had no actors, being an abstract problem in units conversion; the other had an unnecessary gender reference. However, the item was thought to be offensive to Blacks. Perhaps the concern about the item was that it deals with per cents, which are sometimes thought to pose special problems for Blacks. In any case, we do not feel that either item is unfair or offensive.

Only one Paragraph Comprehension item in this book drew more than one comment. This item was the one dealing with urban devlopment in a way that the ETS sensitivity reviewers thought might be offensive to Latin Americans. The comments of seven of the trainees, one of whom was Hispanic, agreed with those of the ETS reviewers. There were four Hispanics who examined this book; since only one of them reacted to it, one might think that the passage does not pose a problem. However, if some of the people from the group to which an item might be offensive react to it, we regard the item as being unfair to them. If a test were run by writing Latin America out of the context, one might find that the rewriting does not affect the item difficulty in the Latin group. It still might be a politically superior strategy to eliminate the Latin reference simply for the sake of appearances.

The group made no comment about the form of the Mechanical Comprehension subtest that was inclosed in the booklet. This subtest contained a fairly androgynous figure to which the item ascribed a sex. The ETS reviewers regarded it as an unnecessary sex reference, but clearly the trainees did not respond to that reference.

Group III

This group consisted of 33 trainees, 24 of whom had no comment. Only two items in this booklet drew more than one comment; both of them were in the Arithmetic Reasoning subtest. One of them had to do with selling a fraction of farm property. No knowledge of farming is required to answer the question, but the trainees tended to feel that such items favored rural folk. Clearly the item is not unfair or offensive. The other item dealt with chances of drawing men and women's names out of hat. Both men and women are mentioned as actors, though names are not given. It is hard to see how the item could be regarded as unfair or offensive.

Group IV

This group consisted of 30 trainees, 21 of whom had no comment on the booklet. Of the Word Knowledge items, only two were cited as unfair. Both items were regarded by a Black as unfair to Blacks; both items were regarded as unfair to "others" by a Black who identified other items as unfair to Blacks. Hence the two responders disagreed as to whom the items were unfair. We do not regard the items as sensitive.

The only other citation by more than one person was in Coding Speed subtest. This was for the item that is similar to "jig" as discussed in the section on the ETS staff review. It was feared that the item would be offensive to Blacks, but none of the six Blacks in this group cited it. It is as if the White and Hispanic trainee that cited the item thought it might be sensitive for Blacks, but the Blacks took an innocuous meaning of the word if they reacted to it at all.

Group V

This group consisted of 42 trainees, 26 of whom had no comment on the booklet. For some reason, perhaps its size, this group made more responses than any other but gave very little description of the problems they observed. Despite the number of items marked, very few of the trainees spoke up during the discussion session that was held at the end after the books were examined. We

will describe the items that were marked most frequently, to indicate the nature of the responses.

The General Science items marked as sensitive included six with no human actors. Hence they were probably not regarded as offensive. They dealt with the solar system, energy sources, earth measurement, insects, and chemistry. Clearly, people with a science background will have an advantage with these items, though why some trainees respond to these and not others is by no means clear. Only one item of this group is recognizeably related to the sensitivity problem and that one deals with skin color, though not in a pejorative way. It was recognized as possibly sensitive by the ETS reviewers as well as the trainees. One Black out of the eleven in this group cited the item as sensitive so it is clearly not blatant. However, the ETS policy would bar its use.

Other General Science items that were marked by several trainees, but not so many as the six mentioned above, were examined. Their citations were for the most part unexplainable. The items appeared to be quite useable from the sensitivity point of view.

Of the Arithmetic Reasoning items in this booklet three were circled by more than one trainee, one circled by three, and two circled by two trainees. One item cited by three trainees had named human actors, both male and female. The item dealt with money, but showed neither sex to a disadvantage. The item marked by two people dealt with a return on investment--the actor was an "investor." It is remotely possible that the third item, which dealt with cooking, might give females some advantage-- all those citing the item were male. All of these items are regarded as useable.

The four Paragraph Comprehension items that were marked by more than one person were items associated with the first stem of the subtest. This stem deals with technical developments in farming, though it is apparent on examination of the item that a knowledge of farming is not required for answering the questions. Still, many of the trainees identified items that might be sensitive to rural-urban differences, and these citations may have occurred for similar reasons. In our instructions to the trainees we used a rural-urban comparison in order to avoid priming the trainees to make particular distinctions when considering other types of groups. This is probably the reason for the citations in this booklet. The items do not appear to be sensitive.

Group VI

Of the forty trainees who examined this booklet 22 had no comment. Only one of the items in the book drew sensitivity citations by more than one trainee, and that one is the Word Knowledge item to which Jewish trainees might be sensitive. As discussed above, that one was comparable to using the word "pogrom" as a stem. Six trainees cited this item. Unfortunately, the only Jewish trainee that took part in this study was in Group I so we cannot see how a Jewish trainee might react to the item. It probably would be easy to find another word to use in place of this item, and we would not recommend its reuse.

REFERENCE

Hunter, R., Slaughter, C. ETS Test Sensitivity Review Process. Princeton, N.J.: Educational Testing Service, 1980.

APPENDIX A: DIRECTIONS FOR THE ETS RESEARCH TEST

HERE IS A NOTE TO THE ETS TEAM TAKING PART IN THE DATA COLLECTION AT FORT DIX ON 14 SEPTEMBER 1983. THE FOLLOWING MATERIAL IS WRITTEN MORE FORMALLY THAN IT PROBABLY SHOULD BE. I DO NOT INTEND FOR THE ADMINISTRATION TO BE STILTED OR FORMAL. BUT I WOULD LIKE YOU ALL TO READ OVER THIS MATERIAL SO THAT YOU CAN SEE WHAT KINDS OF THINGS WE NEED TO DO AND WHAT SOME OF THE CONSIDERATIONS ARE. TIME IS SHORT, I KNOW, BUT I WILL APPRECIATE ANY SUGGESTIONS.

I WISH TO EMPHASIZE THAT THE MATERIAL IS NOT TO BE READ. JUST USE THIS MATERIAL FOR A GENERAL GUIDE.

The trainees will file in and be seated, probably in two rows of approximately 15 for each group. There will be two groups. An administrator will be in front of each group. Each administrator will have one proctor, and there will be one floater for the whole room. When a particular administrator's group is in position, the proctor will distribute to each person in the group a pencil and a booklet that is appropriate for the group. PASS THE BOOKLETS OUT IN CONSECUTIVE NUMERICAL ORDER SO THAT THE BOOK NUMBERS WILL INDICATE THE SEAT LOCATION. NOTE THAT THIS MEANS THAT ONLY THE LOWER NUMBERED BOOKS WILL BE DISTRIBUTED. The booklets will be obtained from the floater before each group enters, as will an additional supply of pencils. After the booklets are passed out, any leftovers will be returned to the floater. Floater will make a record of the books distributed on the appropriate form. Be sure the record of the test forms is maintained completely at this time and no missing forms are noted. If they are, find them.

In the material below, "SAY" is a direction to the administrator.

SAY: "Good Morning! Each of you should have a booklet and a pencil. If you are missing either, please hold up your hand." Missing materials will be distributed to the trainees. Be sure the record of test forms is maintained completely at this time and no missing forms are noted. If they are, find them.

SAY: "As you can see from the front page of the booklet, these are research tests assembled by Educational Testing Service, for whom we work. You will not be taking the the tests in these booklets this morning so don't panic. We only want to find out if the test items are unsuitable from your point of view. We are doing this for the Department of Defense to discover any unnecessary features of the test items that might interfere with the attention you might give to solving the actual mental problems that are supposed to be posed by the test items. We want to be sure that the items do not unfairly penalize any special group of people because they haven't been exposed to some essential piece of information. For example, an item that assumes some special piece of knowledge about the subway would probably unfairly penalize people who had spent their life on the farm. Some items might not be equally fair to people of different sexes, different races, religions, or national origins."

SAY: "Now open your booklets to the first page, and read the directions silently to yourself as I read them aloud."

SAY: "ITEM RATING STUDY DIRECTIONS"[almost all of the material immediately following comes from the directions in the test books].

"The kinds of questions in this booklet are some of the kinds that are used to test abilities for the service. Many studies have shown that these kinds of questions are valid for this use, but some questions may be unfair or offensive to women, Black people, Hispanic people, or some religious groups."

"Please read the questions in this booklet and put a circle around the number of any question you think is unfair or offensive. Afterwards, you will be asked to say why you feel that way about it. Do not try to answer the questions, and do not pay attention to whether the questions are hard or easy, but decide whether the questions are fair and speak of people respectfully."

NOW SAY: "ARE THERE ANY QUESTIONS?"

You should be pretty accepting of questions, but please keep in mind that this is NOT a consciousness raising session. As much as we can we would like the responses of the trainees to be like those they might have made if they were encountering these items WITHOUT any intervention on our part.

NOW SAY: "For our study purposes we are going to ask you some questions about yourselves. But notice that we are not going to ask you your name, and we assure you that your responses on the booklets will not be connected with you by anybody after you leave this room for any reason. Are there any questions about that? "

SAY: "On the page before you please put a circle around the word that correctly indicates your sex."

After a brief pause SAY: "You should now have a circle around either the word 'Male' or the word 'Female.' Please do so if you have not already."

SAY: "Now please circle the word that indicates whether you are Black, Hispanic or Other. You should circle at least one of these three alternatives. Please mark appropriately now." More than one response is allowed for this set only.

NOW SAY: "Finally, please circle the word that describes your religion--Catholic, Jewish, Protestant, or Other. You should circle one of these alternatives. Please mark appropriately." [If they ask, NONE would be marked OTHER].

THEN SAY: "You do NOT need to give your name or other identification."

FINALLY SAY: "We very much appreciate your participation in this study. We hope it will help make test-taking less stressful for those who enter the service in the future. But you should know that you can stop at any time, if you wish, and return to your company. We hope you won't do that, but you may if you must. Any questions?"

Before the beginning of the sessions with trainees, the proctors and administrators were given a set of directions concerning the specific subtests. In these directions one page was devoted to the instructions that went with the particular subtests for each Group. These instructions indicated that examples would be taken from the books handed out to the trainees. This was possible because the trainees were not actually taking the subtests. The instructions indicated the item the administrator should use as an example, how to work it in the case of Arithmetic Reasoning, or why the answer is correct in the case of Mechanical Comprehension. This was done so that the administrator would not be distracted while reading and working the examples. Providing the examples to the administrators was particularly useful for the Arithmetic Reasoning and Paragraph Comprehension subtests because for both of these, if used as an example, the first item or paragraph could involve complications that would distract from the task of understanding what was needed in responding to the items. Since different groups examined different subtests, different examples were needed for each group. Hence a separate set of directions was made up for each group. The directions given for each type of subtest is a short paraphrase of the directions given in the operational ASVAB booklet, and are given below.

GENERAL SCIENCE

SAY: "This test has questions about science. Each examinee picks the BEST answer for each question and marks it on the answer sheet."

ARITHMETIC REASONING

SAY: "This is a test of arithmetic word problems. Each question is followed by four possible answers. The examinee decides which answer is CORRECT, then, on the answer sheet, indicates the letter that is the same as the letter chosen as CORRECT."

WORD KNOWLEDGE

SAY: "This is a test of the examinee's knowledge of word meanings. The questions consist of a sentence of phrase with a word or phrase underlined. From the four choices given, the examinee decides which ONE MEANS THE SAME, OR MOST NEARLY THE SAME, as the underlined ord or phrase. Having made the decision, the examinee marks the answer on the answer sheet."

PARAGRAPH COMPREHENSION

SAY: "This is a test of the examinee's ability to read with understanding. One or more paragraphs of reading material are followed by incomplete statements or questions. The examinee reads the material and selects one of four lettered choices that BEST completes the statement or answers the question. The selection is then marked on the answer sheet."

CODING SPEED

SAY: "This is a test to see how quickly and accurately the examinee can find a number in a table. At the top of each page there is a number table or 'key.' The key is a group of words with a code number for each word. Each question in the test is a word taken from the key at the top of that page. From among the possible answers listed for each question the examinee finds the one that is the correct code number for that word, and then blackens the corresponding place on the answer sheet."

MECHANICAL COMPREHENSION

SAY: "This test has questions about mechanical and physical principles. The examinee studies the pictures and decides which answer is correct, then marks the appropriate space on the answer sheet."

* * * * * * * * *

After reading one of the short paragraphs of direction the administrator would read the item stem and alternatives, and then indicate the correct answer, why it is correct, and that the examinee would then record the answer on an answer sheet. Since all the trainees had taken the ASVAB, they had encountered that procedure.

Each group was taken through the directions for all three of the subtests that they would examine. Then they were told to begin as can be seen in the script below.

THEN SAY: "Remember, for each item that you think is unfair or offensive for some group, please put a circle around the item number. Also, so you will remember, write the name of the group for whom you think the item is unsuited and, if you can, circle the part that causes the trouble. Are there any questions? Then go to the beginning of the first test and begin. Remember, circle the question with the problem, indicate for which group or groups there is a problem and, if you can, circle the part of the item that causes the trouble."

The proctor and the administrator can watch the trainees, answer any questions, be sure the circled items have groups named and circled parts if the trainees can give these indications. Periodically the administrator can remind the trainees what they are supposed to be doing. Also periodically ask how many have finished. Since we have to have the trainees out of the room at the end of the hour, we have to start discussions well before that. When the administrator feels that a reasonable amount of time has passed, he should follow the procedure given below.

SAY: "Now let's turn to the first page of the first test. Has anyone circled any items?"

Then pick someone to identify the item they have circled and why. Make sure it is marked on the booklet. Both the administrator and the proctor should make notes on every comment made aloud. Go through the whole page taking everyone's comment. If you have had one comment on an item, ask if there are other comments on the item. Keep making notes and encourage the trainees to indicate the group and the troublesome part---BUT ONLY IF THEY HAD CIRCLED THE ITEM BEFORE THE PUBLIC REVIEW STARTED. [THE POINT HERE IS THAT WE DON'T WANT THE TRAINEES TO PUT IN THEIR INDIVIDUAL BOOKS THE COMMENTS THAT THEY'VE HEARD OTHERS MAKE]. Procede page by page until you are through the book.

Then SAY: "Now we are finished. I hope you have told us what you think we need to know about these test items. Both ETS and the Department of Defense appreciate your work here this morning. NAME OF THE PROCTOR will come around and pick up the booklets and pencils."

The proctor should now pick up the books and make very sure that there is one book returned for each examinee. THE BOOK NUMBERS CAN BE USED TO INDICATE THE SEAT WHERE ANY MISSING BOOKLET WAS USED.

SAY: "Now you may leave."